④

SCALING PROPERTIES OF
COARSE-CODED SYMBOL MEMORIES

Ronald Rosenfeld and David S. Touretzky

Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

The Artificial Intelligence
and Psychology Project

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

88 12 28 051

SCALING PROPERTIES OF
COARSE-CODED SYMBOL MEMORIES

Technical Report AIP-10

Ronald Rosenfeld and David S. Touretzky

Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA 15213

29 September 1987

ADA204456

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
Unclassified	

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
	Approved for public release;
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
A I P - 10	

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Carnegie-Mellon University		Computer Sciences Division
		Office of Naval Research (Code 1133)

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Department of Psychology	800 N. Quincy Street
Pittsburgh, Pennsylvania 15213	Arlington, Virginia 22217-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Same as Monitoring Organization		N00014-86-K-0678

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS p400005ub201/7-4-86			
	PROGRAM ELEMENT NO	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO
	N/A	N/A	N/A	N/A

11. TITLE (Include Security Classification)
Scaling Properties of Coarse-Coded Symbol Memories

12. PERSONAL AUTHOR(S)
R. Rosenfeld and D.S. Touretzky

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	87 September 29	4

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Articial Intelligence, Machine Learning, Connectionism,
			Short Term Memory, distributed representation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Coarse coded memories have appeared in several neural network symbol processing models, such as Touretzky and Hinton's distributed connectionist production system DCPS, Touretzky's distributed implementation of Lisp S-expressions on a Boltzmann machine, and St. John and McClelland's PDP model of case role defaults. In order to determine how these models would scale, one must first have some understanding of the mathematics of coarse coded representations. For example, the working memory of DCPS, which stores triples of symbols and consists of 2,000 units, can hold roughly 20 items at a time out of a 15,625-symbol alphabet. How would DCPS scale if the alphabet size were raised to 50,000? With the current alphabet size, how many units would have to be added simply to double the working memory capacity to 40 triples? We present some analytical results related to these questions. (KF)

The authors

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Alan L. Meyrowitz	(202) 696-4302	N00014

DD FORM 1473, 84 MAR
83 APR edition may be used until exhausted.
All other editions are obsolete.

Scaling Properties of Coarse-Coded Symbol Memories

Ronald Rosenfeld

David S. Touretzky

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Summary

Coarse coded memories have appeared in several neural network symbol processing models, such as
Touretzky and Hinton's distributed connectionist production system DCPS [6,7], Touretzky's distributed
implementation of Lisp S-expressions on a Boltzmann machine, BoltzCONS [8,9], and St. John and
McClelland's PDP model of case role defaults [4]. In order to determine how these models would scale,
one must first have some understanding of the mathematics of coarse coded representations. For example,
the working memory of DCPS, which stores triples of symbols, consists of 2,000 units, can hold roughly
20 items at a time out of a 15,625-symbol alphabet. How would DCPS scale if the alphabet size were
raised to 50,000? With the current alphabet size, how many units would have to be added simply to
double the working memory capacity to 40 triples? We present some analytical results related to these
questions.

A coarse coded symbol memory in its most general form is defined by two parameters: the alphabet
size α and the number of units, N. Each unit has a "receptive field" containing some subset of the
alphabet. Symbols are stored in memory by turning on all the units in whose receptive field they fall.
Thus, symbols are represented as distributed patterns of activity, and the units are said to be "coarsely
tuned" because each participates in the representation of more than one symbol. However, our units'
receptive fields are not restricted to contiguous subregions of a multidimensional feature space as are
the "value units" of [1,2,3,5]. They are instead random subsets of a one-dimensional symbol space.

A symbol is deemed present if all its receptors are active (our analysis easily generalizes to a weaker
criterion). As items are added, "ghost" symbols eventually appear; these are symbols which were not
stored, but appear because all their receptors are shared with symbols that were stored. The capacity
or "span" of a memory is the number of symbols k that can be stored before ghosts appear. (A localist
representation, where $k = \alpha = N$, is very inefficient for sparse memories with a large alphabet.)

In this analysis we assume that symbols have a uniform receptor set size L, and that each of the

1

α symbols is assigned a random L-subset of the N units making up the memory. The probability of a ghost appearing after k symbols have been stored is given by Equation 1:

$$P_{\text{ghost}}(N, L, k, \alpha) \quad = \quad 1 - \sum_{c=0}^{N} T_{N,L}(k, c) \cdot \left[1 - \frac{\binom{C}{L}}{\binom{N}{L}}\right]^{\alpha - k} \tag{1}$$

$T_{N,L}(k, c)$ is the probability that exactly c units will be active after k symbols have been stored. It is defined by Equation 2:

$$T_{N,L}(k, c) \quad = \quad \sum_{a=0}^{L} T(k-1, c-a) \cdot \frac{\binom{N-(c-a)}{a}\binom{c-a}{L-a}}{\binom{N}{L}} \tag{2}$$

$$T_{N,L}(0, c) = 0 \quad \text{for all } c.$$

The optimal pattern size with respect to N, k, and α can be determined by binary search on Equation 1. However, this may be expensive for large N. A good initial estimate is the L that maximizes the following expression:

$$\frac{\binom{N \cdot \left[1 - (1 - L/N)^k\right]}{L}}{\binom{N}{L}} \tag{3}$$

We have constructed coarse coded memories of various sizes and measured their capacities experimentally. The results show good agreement with the predicted values.

We present graphs of the relationships between N, k, α, and P_{ghost} for optimum pattern sizes, as determined by Equation 1. A representative graph is shown in Figure 1. The results show an exponential relationship between α and N/k. Thus, for a fixed alphabet size, the span is proportional to the number of units. For $P_{\text{ghost}} = 0.01$ the relationship is:

$$\alpha \quad = \quad e^{\left[0.468\frac{N}{k} - 4.76\right]} \tag{4}$$

We compare the capacity obtained using our probabilistic, random receptive fields approach with that of two other approaches which guarantee a specified span: a binary coding scheme, and an approach where the overlap between any two patterns is bounded by $\lfloor (L-1)/k \rfloor$.

References

[1] Ballard, D H. (1986) Cortical connections and parallel processing: structure and function. *Behavior and Brain Sciences* 9(1).

[2] Feldman, J. A., and Ballard, D. H. (1982) Connectionist models and their properties. *Cognitive Science* 6, pp. 205-254.

[3] Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. (1986) Distributed representations. In D. E. Rumelhart and J. L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. Cambridge, MA: MIT Press.

[4] St. John, M. F. and McClelland, J. L. (1986) Reconstructive memory for sentences: a PDP approach. *Proceedings of the Ohio University Inference Conference.*

[5] Sullins, J. (1985) Value cell encoding strategies. Technical report TR-165, Computer Science Department, University of Rochester, Rochester, NY.

[6] Touretzky, D. S., and Hinton, G. E. (1985) Symbols among the neurons: details of a connectionist inference architecture. *Proceedings of IJCAI-85*, Los Angeles, CA, pp. 238-243.

[7] Touretzky, D. S., and Hinton, G. E. (1986) A distributed connectionist production system. Technical report CMU-CS-86-172, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA.

[8] Touretzky, D. S. (1986) BoltzCONS: reconciling connectionism with the recursive nature of stacks and trees. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, Amherst, MA, pp. 522-530.

[9] Touretzky, D. S. (1986) Representing and transforming recursive objects in a neural network, or "Trees *do* grow on Boltzmann machines." *Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics*, Atlanta, GA, pp. 12-16.
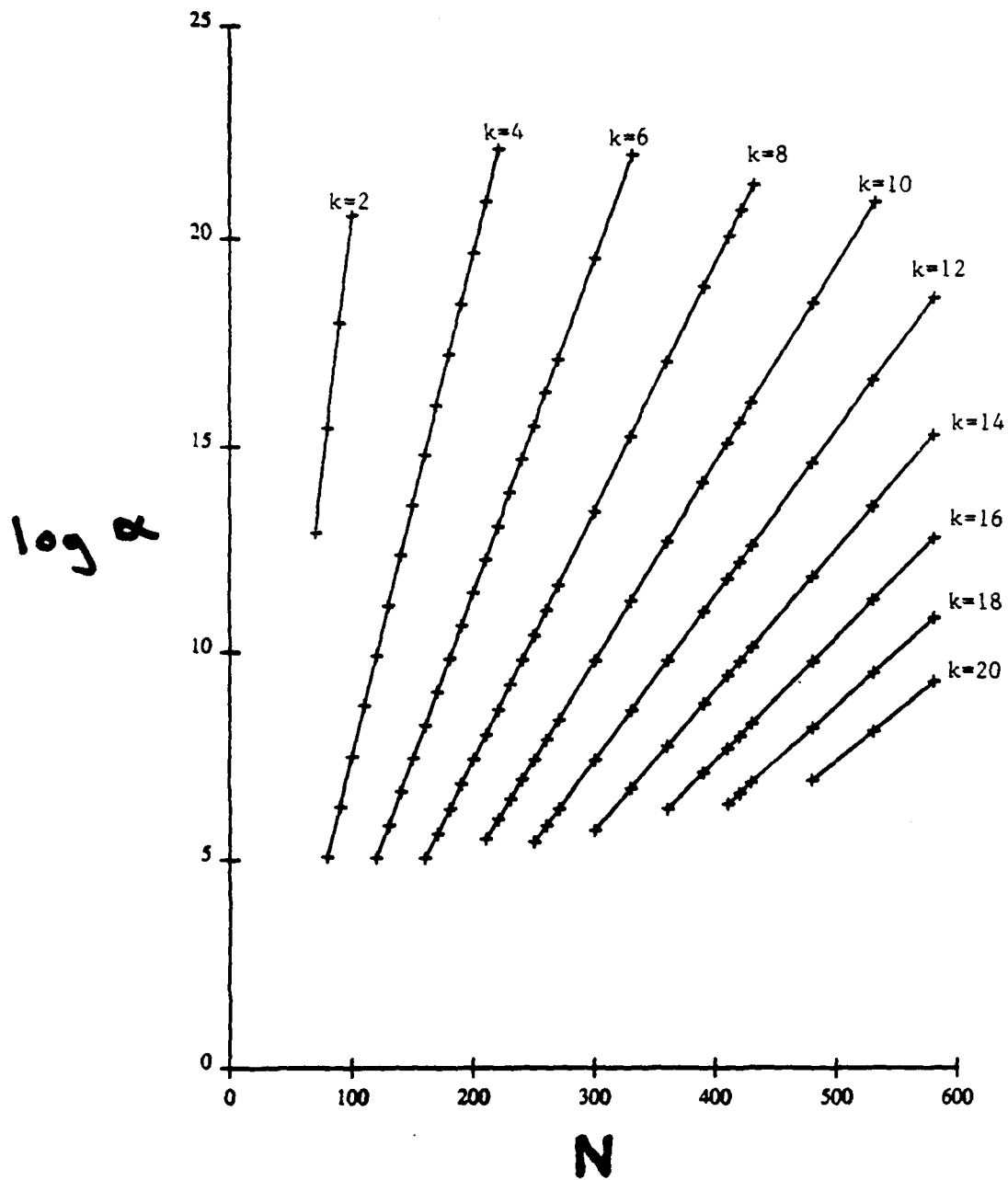
Figure 1: A graph of log α vs. N for even k values (memory capacity) from 2 to 20. $p = 0.01$.